
Data Analysis Project

Data Analyst: Mayuresh Koli

Purpose:

The Marvel Cinematic Universe (MCU) is an American media franchise and shared universe centred on a series of superhero films produced by Marvel Studios. The MCU is expanding each year, and we know how fans wait for each movie with patience and excitement. Each MCU movie earns millions of dollars at the box office worldwide. Domestic box office will always be higher in most cases, but other than that, which country or region is more profitable for the MCU? I am conducting this data analysis to find out the answer.

Dataset Description:

There are six datasets as follows:

1) movie_info dataset -> This dataset includes metadata about the movies like their title, release date, budget, etc. It also includes ratings from different critics.

Column Name	Column Description
<i>movie_title</i>	<i>The title of the movie.</i>
<i>release_date</i>	<i>The release date of the movie.</i>
<i>season</i>	<i>Because the movies are from the United States, the seasons are based on the North American calendar.</i>
<i>phase</i>	<i>The phase of the movie. Marvel phases are a way of grouping several films in the MCU together.</i>
<i>production_budget_in_million_(USD)</i>	<i>The production budget of a movie.</i>
<i>worldwide_collection_in_million_(USD)</i>	<i>World-wide collection of a film.</i>
<i>tomatometer</i>	<i>Ratings or scores from Rotten Tomatoes Critics (Converted into decimal form)</i>
<i>tomato_audience_score</i>	<i>Ratings or scores from Rotten Tomatoes Audience (Converted into decimal form)</i>
<i>imdb</i>	<i>Ratings or scores from IMDB</i>

<i>metascore</i>	<i>Ratings or scores from Metacritic Critics (Converted into decimal form)</i>
<i>meta_user_score</i>	<i>Ratings or scores from Metacritic Users</i>

2) *europa_box_office* -> This dataset includes box office collections from various countries in European region.

Column Name	Column Description
<i>movie_title</i>	<i>The title of the movie.</i>
<i>Country_names</i>	<i>This includes a number of countries from the European region. It includes 29 different countries' columns.</i>
<i>Other_European_Countries</i>	<i>This is a specific column in the <i>europa_box_office</i> dataset that states other European countries that are not included because it is an added column to handle the missing data for the difference between worldwide collection and region-wise collection. There is also no official information about the box office of the other remaining countries.</i>

3) *middle_east_and_africa_box_office* -> This dataset includes box office collections from various countries in Middle east and African region.

Column Name	Column Description
<i>movie_title</i>	<i>The title of the movie.</i>
<i>Country_names</i>	<i>This includes a number of countries from the Middle East and African region. It includes 11 different countries' columns.</i>
<i>Other_Middle_East_and_African_Countries</i>	<i>This is a specific column in the <i>middle_east_and_africa_box_office</i> dataset that states other Middle East and African countries that are not included because it is an added column to handle the missing data for the difference between worldwide collection and region-wise collection. There is also no official information about the box office of the other remaining countries.</i>

4) *asia_pacific_box_office* -> This dataset includes box office collections from various countries in Asia-Pacific region.

Column Name	Column Description
<i>movie_title</i>	<i>The title of the movie.</i>
<i>Country_names</i>	<i>This includes a number of countries from the Asia-Pacific region. It includes 15 different countries' columns.</i>
<i>Other_Asia_Pacific_Countries</i>	<i>This is a specific column in the <i>asia_pacific_box_office</i> dataset that states other Asia-Pacific countries that are not included because it is an added column to handle the missing data for the difference between worldwide collection and region-wise collection. There is also no official information about the box office of the other remaining countries.</i>

5) *north_america_box_office* -> This dataset includes box office collections from various countries in North American Region

Column Name	Column Description
<i>movie_title</i>	<i>The title of the movie.</i>
<i>USA_and_Canada</i>	<i>USA and Canada box office collection, it can also be called as domestic box office collection.</i>
<i>Mexico</i>	<i>Mexico box office collection. It is also a north American country.</i>
<i>Central America</i>	<i>Central America box office collection. it is created to handle the missing data for the difference between worldwide collection and region-wise collection. There is also no official information about the box office of the other remaining countries.</i>
<i>Caribbean</i>	<i>Caribbean box office collection. it is created to handle the missing data for the difference between worldwide collection and region-wise collection. There is also no official information about the box office of the other remaining countries.</i>

6) *south_america_box_office* ->

Column Name	Column Description
<i>movie_title</i>	<i>The title of the movie.</i>

<i>Country_names</i>	<i>This includes a number of countries from the Asia-Pacific region. It includes 10 different countries' columns.</i>
<i>Other_South_American_Countries</i>	<i>This is a specific column in the south_america_box_office dataset that states other South American countries that are not included because it is an added column to handle the missing data for the difference between worldwide collection and region-wise collection. There is also no official information about the box office of the other remaining countries.</i>

Important Notes:

- *Because the MCU began in 2008, it will not include any Marvel films that are not part of the MCU.*
- *Marvel TV series and animated series are not included.*
- *The dataset will be updated regularly once a new movie is released and their box office is published.*
- *Due to no official information about some countries' box office collections, there can be null values. There are many types to deal with and it may depend on your analysis.*
- *Another reason for null values is that some movies were never released in some countries, perhaps due to their censor board or may be due to pandemic.*

Data Extracted From:

- https://www.boxofficemojo.com/?ref=bo_nb_hm_mojologo
- https://en.wikipedia.org/wiki/List_of_Marvel_Cinematic_Universe_films
- <https://www.metacritic.com/>
- <https://www.imdb.com/>
- <https://www.rottentomatoes.com/>

Tableau Visualization link:

<https://public.tableau.com/app/profile/mayuresh.koli/viz/MarvelCinematicUniverseWorldwideBoxOfficeCollection/Dashboard1>

My Profiles links:

- Github -> <https://github.com/MayureshKoli45>
- Kaggle -> <https://www.kaggle.com/mayureshkoli>
- Tableau -> <https://public.tableau.com/app/profile/mayuresh.koli#!/>
- LinkedIn -> <https://www.linkedin.com/in/mayuresh45/>

